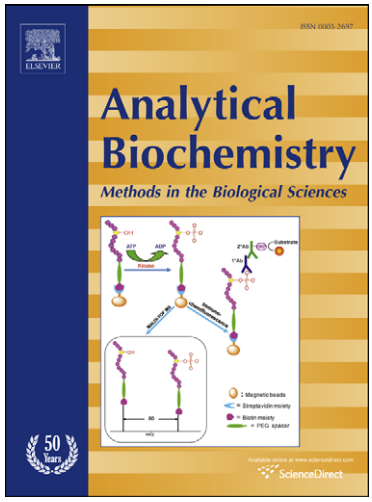
A plasmid-based lacZα gene assay for DNA polymerase fidelity measurement

Article (Accepted Version)



A plasmid-based *lacZα* gene assay for DNA polymerase fidelity measurement

Brian J. Keith, Stanislaw K. Jozwiakowski, Bernard A. Connolly

PII: S0003-2697(12)00530-1
DOI: http://dx.doi.org/10.1016/j.ab.2012.10.019
Reference: YABIO 11127

To appear in: *Analytical Biochemistry*

Received Date: 31 August 2012
Revised Date: 10 October 2012
Accepted Date: 12 October 2012

Please cite this article as: B.J. Keith, S.K. Jozwiakowski, B.A. Connolly, A plasmid-based *lacZα* gene assay for DNA polymerase fidelity measurement, *Analytical Biochemistry* (2012), doi: http://dx.doi.org/10.1016/j.ab.2012.10.019

A plasmid-based *lacZα* gene assay for DNA polymerase fidelity measurement

Brian J. Keith[1], Stanislaw K. Jozwiakowski[2] and Bernard A. Connolly[1]*

Institute of Cell and Molecular Biosciences (ICaMB), Newcastle University, Newcastle upon Tyne, NE2 4HH, UK

[1]Newcastle University, address as above

[2]MRC Genome Damage and Stability Centre, University of Sussex, Brighton, BN1 9RQ, UK

***To whom correspondence should be addressed. Tel: +44 0191-222 7371; Fax: +44 0191 222 7424; Email: bernard.connolly@ncl.ac.uk**

Running Head: Measuring DNA polymerase fidelity

Subject category: Enzymatic assays and analysis

ABSTRACT

A significantly improved DNA polymerase fidelity assay, based on a gapped plasmid containing the *lacZα* reporter gene in a single-stranded region, is described. Nicking at two sites flanking *lacZα,* and removing the excised strand by thermocycling in the presence of complementary competitor DNA, is used to generate the gap. Simple methods are presented for preparing the single-stranded competitor. The gapped plasmid can be purified, in high amounts and in a very pure state, using benzoylated-naphthoylated-DEAE-cellulose, resulting in a low background mutation frequency (~ $1x10^{-4}$). Two key parameters, the number of detectable sites and the expression frequency, necessary for measuring polymerase error rates have been determined. DNA polymerase fidelity is measured by gap filling, *in vitro*, followed by transformation into *E. coli* and scoring of blue/white colonies and converting the ratio to error rate. Several DNA polymerases have been used to fully validate this straightforward and highly sensitive system.

Keywords: DNA polymerase; polymerase fidelity; lacZα reporter gene

INTRODUCTORY STATEMENT

DNA replication is a complex process that requires a multi-protein assembly (the replisome) to bring about the efficient, rapid and accurate copying of the genome [1,2]. The replisome minimises the appearance of detrimental mutations, thereby maintaining genomic stability, and the base substitution error rate of the *E. coli* replication machinery *in vivo* is about $1x10^{-7}$-10^{-8} [3]. DNA polymerases make a major contribution to accuracy and some replicative polymerase result in only one error per $10^5/10^6$ bases incorporated, although other enzymes e.g. those involved in translesion bypass are less accurate [4,5]. DNA polymerases are essential components of many in biotechnological applications, especially the polymerase chain reaction (PCR) where accuracy is important [6-8]. Therefore, it is important both to understand the mechanisms that contribute to DNA polymerase fidelity and to develop assays sensitive enough to detect the infrequent errors that many polymerases make.

Several approaches have been described to measure fidelity; however many are complicated or do not detect the whole spectra of mutations that may occur [9-11]. One of the most successful methods uses replication of a gap in the bacteriophage M13mp2 *lacZα* gene, encoding an inactive fragment of β-galactosidase, the α-peptide. When accurately copied, and subsequently introduced into *E.coli* that bears a complementing copy of the remaining β-galactosidase gene, functional β-galactosidase is reconstituted, resulting in the hydrolysis of X-gal and blue bacterial plaques [12]. Inaccurate polymerase activity may result in a defective α-peptide, eventually resulting in reduced or abolished β-galactosidase activity, indicated by light-blue or colourless plaques. The error rate is calculated from the blue/colourless plaque ratio, and further information can be obtained from DNA sequencing. This method allows the detection of all twelve possible base substitutions as well as insertions and deletions in varying sequence context, is very well characterised and has been extensively used by the Kunkel group for many investigations of DNA polymerases [4,5,12,13].

Previously we described a plasmid-based method for measuring polymerase fidelity using pSJ1, a derivative of pUC18 [14]. In pSJ1 a segment encoding *lacZα* is flanked by two single-strand nicking endonuclease sites, allowing one of the DNA duplex strands to be specifically cut on both sides of the *lacZα* gene. Removal of the nicked DNA fragment, results in a gapped plasmid containing the *lacZα* gene in the single-stranded region. The accuracy of a polymerase can be determined by copying the gene *in vitro* and then introducing the plasmid into *E.coli*, an approach similar in concept to that described above for the bacteriophage system. The plasmid-based system benefits from the simple preparation of pSJ1 in reasonable quantities and straightforward use in fidelity evaluation [14]. As plasmids are compatible with many cell types, gapped variants can be used to study DNA replication and repair *in vivo* [15]. Nevertheless, the pSJ1 method is presently underdeveloped as compared to the much longer established bacteriophage approach. In particular two key parameters, the number of detectable mutations (alterations in *lacZα* which result in an inactive gene and a white phenotype) and the expression frequency (the degree to which the polymerase-synthesised strand is expressed in *E.coli*) have yet to be determined. Knowledge of these features is required for converting the observed ratio of blue/white colonies/plaques to a polymerase error rate, measured as mistakes per base incorporated [12,13]. In this publication both the detectable mutations and the expression frequency have been measured, making the plasmid system fully equivalent to the bacteriophage. Further, considerable improvements in the preparation and purification of the gapped plasmids are outlined, which result in larger amounts of purer product and lowered background mutation rates. Together these refinements substantially increase the scope and power of plasmid methods for polymerase fidelity measurements.

MATERIALS AND METHODS

Materials

Deoxynucleotide triphosphates (dNTPs), lambda (λ) exonuclease, ExoIII, DNA ligase and the restriction endonucleases EcoRI, DpnI, NdeI and SalI were supplied by Fermantas (York, UK). Taq-Pol, for use in routine PCR applications was also purchased from Fermantas. The nicking enzymes Nt.Bpu10I, Nt.BbvCI and the restriction enzyme BbvCI were purchased from New England Biolabs (Ipswich, USA). Plasmid safe DNase was supplied by Cambio (Cambridge, UK). PCR clean-up, nucleotide removal, miniprep and maxiprep kits were from Qiagen (Crawley, UK), Benzoylated-naphthoylated-DEAE-cellulose (BND-cellulose) was from Sigma-Aldrich (Gillingham, UK). Top10 competent *E. coli* were from Invitrogen (Paisley, UK) and Velocity DNA polymerase from Bioline (London, UK). *Pyrococcus furiosus* family B DNA polymerase (Pfu-Pol B), Pfu-Pol B exo$^-$ (D141A/E143A), Pfu-Pol B exo$^-$ error prone (D141A/E143A/D473G) was purified as previously described [16,17]. A Taq-Pol overexpressing system was created by inserting the Taq-Pol gene (amplified by PCR from *Thermus aquaticus* genomic DNA) between the NdeI and SalI restriction endonuclease sites of pET-17b (Novagen, Merck-Millipore, Watford, UK) using DNA ligase. The resulting pET-17b(Taq-Pol) was used to transform *E. coli* BL21(DE3) (pLysS) and the polymerase purified exactly as described for Pfu-Pol [16,17]. Image quant colony counting software was provided by GE Healthcare (Hatfield, UK).

Preparation of pSJ2 and pSJ3

Two plasmids pSJ2 and pSJ3, based on pUC18 and M13mp2, were produced for DNA polymerase fidelity assays. pSJ2 was prepared and supplied by Biomatik (Delaware, USA). pSJ3 was assembled in house commencing from pUC18. A PCR-based site directed mutagenesis protocol [18] was used to: 1) flank the *lacZα* gene with N(t/b)Bpu10I nicking sites (equivalent to a Bpu10I restriction site); 2) delete the dam methylase (GATC) site within this gene, in order to avoid cellular mismatch-directed repair mechanisms [19]. Reactions were performed in a 100 µl reaction volume and consisted of 200 ng template DNA, 1.5 µM

of each primer, 400 µM of each dNTP, 4 units of Velocity and supplier's recommended buffer. The reaction was subjected to 32 PCR cycles of 35 s at 95°C, 35 s at 55°C and 3 minutes 30 s at 70°C. An initial denaturing step of 98°C for 1 min and a final 10 min elongation step were also performed. The polylinker in pUC18 was deleted using the PCR-based SLIM (site-directed ligase independent mutagenesis) approach with appropriately designed primers [20]. Two individual reactions, each half the volume described above, were submitted to 22 PCR cycles. These reactions were then combined to form a new reaction and subjected to a further 22 PCR cycles after the addition of a further 2 units of Velocity. Each PCR reaction used the same initial, extension and final cycles described above. After all mutagenesis PCR reactions the resulting plasmid was purified using a PCR clean up kit according to the supplier's protocol, with the product eluted in 35 µl of H_2O preheated to 50 °C. 15 µl of this product was supplemented with 7 units of DpnI in the supplier's recommended buffer and incubated at 37 °C for 3 hours to degrade the parental plasmid. 1 µl of this reaction was used to transform 60 µl of competent *E. coli* Top10 cells and recovered plasmids were completely sequenced to confirm the integrity of the mutagenesis.

Identifying detectable sites in pSJ2 and pSJ3

The *lacZα* DNA sequences in both pSJ2 and pSJ3 are very similar to that found in M13mp2. Therefore, almost all the detectable sites (i.e. base changes that result in an inactive α-peptide) are known. The status of the small number of *lacZα* bases in pSJ2 and 3, not found in the bacteriophage, was determined using site-directed mutagenesis. The deoxynucleoside at each of these uncharacterised positions was separately changed to all other three bases (for the promoter region) and to all bases that resulted in a new amino acid (for the protein coding region). PCR-based site-directed mutagenesis [18] was used, as described above, to make these alterations.

Preparation of single-stranded *lacZα* competitor DNA

Single-stranded DNA was generated using a PCR/λ-exonuclease method [14]. Two primers, one containing a 5′-phosphate and the second three phosphorothioates at the phosphodiesters nearest the 5′ end (supplementary material, figure S1) were used to amplify the *lacZα* gene in pSJ2 and pSJ3. The PCR reaction mixture (100 µl) consisted of: plasmid (300 ng), primers (1.5 µM of each), dNTPs (200 µM of each), $MgCl_2$ (1.5 mM), Taq-Pol (5 units as defined by the supplier) in the supplier's recommended buffer. PCR consisted of 30 cycles: 35 s at 95 °C (1 minute at 98 °C for the first cycle), 35 s at 55 °C, 30 s (90 s on last cycle) at 70 °C. The amplified *lacZα* DNA was purified with a PCR clean-up kit and the 5′-phosphorylated strand specifically degraded. 2 µg of amplified product was treated with 5 units of λ-exonuclease in 50 µl of the supplier's recommended buffer for 30 minutes at 37 °C. The resulting single-stranded DNA was purified using a nucleotide removal kit. Alternatively chemically synthesised oligodeoxynucleotides were used as competitors (supplementary material, figure S1).

Preparation of gapped DNA

40 µg of plasmid was digested with 40 units of Nt.BbvCI (pSJ2) or Nt.Bpu10I (pSJ3), both enzymes cut twice on the coding strand flanking the *lacZα* gene. Nicking was performed for 3 hours at 37 °C in a 1 ml of the supplier's recommended buffer. The nicked plasmids were subsequently purified using a PCR clean-up kit. This reaction was carried out five times in parallel (total of 200 µg plasmid). Nicked plasmid was gapped in forty eight simultaneous reactions each containing 3.5 µg (about 2 pmol) of plasmid using either a 10-fold molar excess of PCR/λ-exonuclease-prepared single-stranded competitor (for pSJ2) or a 50-fold molar excess of chemically synthesised oligodeoxynucleotide competitor (for pSJ3). The aliquots of nicked plasmid and competitor DNA were each incubated in 100 µl of the nicking buffer and subjected to 3 heat/cool cycles of 95 °C for 1 minute, 60 °C for 10 minutes, 37 °C

for 20 minutes. The reaction mixtures were then pooled and gapped plasmid was separated from excess competitor and any displaced double-stranded *lacZα* DNA using four passes (reducing the volume ~ 10-fold in each pass) through a 100 kDa cut-off Amicon ultrafilter (Millipore, Watford, UK). During ultrafiltration the gapped plasmid was simultaneously equilibrated to 10 mM Tris-HCl (pH 8.0), 1 mM EDTA (TE buffer). The process yielded 160 µg of crude gapped plasmid which was subsequently purified using 2 g of BND-cellulose prepared as a 50 % suspension in TE buffer containing 1 M NaCl. The DNA was gently agitated with 4 ml of BND-cellulose slurry for 10 minutes at room temperature and the mixture loaded into a column. The resin was washed with 60 ml TE buffer containing 1.0 M NaCl and the gapped plasmid eluted with CFS buffer (TE buffer containing 1 M NaCl, 2 % (w/v) caffeine and 50% (v/v) formamide). 1 ml fractions were collected and assayed for the presence of gapped plasmid using 1% agarose gel electrophoresis. Appropriate fractions were pooled and the gapped plasmid re-equilibrated to TE buffer using ultrafiltration as described above. Purified gapped plasmid (80 µg) was stored frozen in 20 µl aliquots at -20 °C.

Expression frequency determination

This investigation was only carried out with pSJ2. Initially PCR-based site-directed mutagenesis was used to destroy the upstream nicking site, giving a derivative (pSJ2A) with only a single downstream nicking site. A further round of mutagenesis with pSJ2A was used to make a solitary base change, which resulted in an in-phase premature stop codon towards the beginning of the *lacZα* coding sequence, yielding pSJ2B. The two plasmids, pSJ2A and B, were then used to produce a heteroduplex with a single base mismatch [15]. To achieve this, pSJ2A (40 µg) was nicked on one strand at its single Nt.BbvCI site (40 units of enzyme for 16 hours at 37 °C) and the cut strand degraded with ExoIII (2400 units of ExoIII for 20 minutes at 37 °C in 1 ml of the supplier's recommended buffer). The resulting single-stranded circular DNA was purified using a PCR clean-up kit. pSJ2B (10 µg) was cut on both strands at its single BbvCI site by digestion with 40 units of the restriction endonuclease BbvCI (16 hours

at 37 °C) and the resulting linear duplex purified with a PCR clean-up kit. The cutting/purification cycle with pSJ2B was repeated twice to ensure complete digestion. The single-stranded circular DNA (derived from pSJ2A) and linear duplex (derived from pSJ2B) were mixed in a 1:1.5 molar ratio, heated to 95 °C and cooled slowly to room temperature to produce the required circular heteroduplex DNA, containing a single nick. The resulting solution was treated with plasmid-safe DNase, which destroys any remaining single-stranded circular DNA and linear duplex DNA, but does not degrade the heteroduplex. 10 units of plasmid-safe DNase were used in the supplier's recommended buffer supplemented with 1 mM ATP. After treatment for 15 hours at 37 °C the heteroduplex was purified with a PCR clean-up kit. As above the DNaseI/purification cycle was repeated twice more to maximise degradation of unwanted starting components. To produce an appropriate control this process was repeated but both the single-stranded circular DNA and the liner duplex were derived from pSJ2A, ultimately resulting in a wild type functional *lacZα* gene. The heteroduplex and the control were used to transform *E. coli* Top 10 (see below) and the ratio of blue/white colonies determined, allowing calculation of the expression frequency.

Background mutation rate determination

The background mutation frequency was determined by transforming 60 µl of *E. coli* Top 10 cells [14] with 1 µl of the plasmid under investigation (pSJ2, pSJ3 and their gapped derivatives). Following transformation these cells are then plated on LB agar supplemented with 40 µg/ml X-Gal, 100 µg/ml ampicillin and 250 µM IPTG. These plates are incubated at 37°C for 16 hours and then scored for blue and white colonies.

DNA polymerase fidelity assay

The gapped derivatives of pSJ2 and pSJ3 were filled by the DNA polymerase under investigation using 20 µl of 20mM Tris-HCl (pH 8.0), 10 mM KCl, 2 mM $MgSO_4$, 10mM $(NH_4)_2SO_4$, 0.1% Triton X-100, 2 µg bovine serum albumin and 250 µM of each dNTP. The

concentrations of the DNA polymerase and the gapped plasmid were 100 nM and 1 nM, respectively. Extension reactions were carried out at 70 °C for 30 minutes and, after this time, examined using 1% agarose gel electrophoresis. About 9 µl of the reaction mixture was analysed directly by electrophoresis and a further 9 µl was tested following digestion with 5 units of EcoRI (30 minutes at 37 °C). Providing the electrophoresis/EcoRI analyses indication complete extension, 1 µl of the reaction mixture was used for the transformation of competent *E. coli* Top 10 cells (60 µl). Transformed cells were plated on LB agar and scored for blue/white colonies as described above. In order to determine the appropriate background mutation rate for polymerase extension assays, the protocol was carried out in the absence of DNA polymerase (in this case electrophoretic analysis is not required).

Colony counting

In order to facilitate counting large numbers of blue and white *E. coli* colonies, a digital camera was used to take images of the plates, which were then analysed by Image Quant colony counting software.

DNA sequencing

Mutant (white) colonies were grown overnight at 37°C in LB media containing 100 µg/ml ampicillin. Plasmid was purified using miniprep kits and sent to be sequenced. (GATC-Biotech, Cambridge, UK).

RESULTS

Redesigning the *lacZα* gene to give pSJ2 and pSJ3

In order to rigorously evaluate the error rate of a polymerase as mistakes made per base incorporated, it is necessary to know the number of detectable mutations in the indicator gene. A particular advantage of the bacteriophage system, based on the *lacZα* sequence present in M13mp2, is that a large number of experiments have characterised almost every detectable

mutation i.e. exactly which base substitutions and insertions/deletions result in an inactive α-peptide and hence a change in plaque colour from blue to white [4, 5, 12, 13, 21, 22]. The plasmid previously used pSJ1, encodes an active α-peptide, but its gene sequence is derived from pUC18 and, ultimately, M13mp18 [14]. Thus, as shown in figure 1A, the M13mp2 and pSJ1 *lacZα* sequences vary somewhat, making it complicated to apply the wealth of knowledge available for the phage system to this plasmid. As a remedy two new plasmids have been designed with *lacZα* sequences that more closely match that found in M13mp2. One of the plasmids, pSJ2, has a *lacZα* sequence almost identical to M13mp2, mainly differing in the presence of three and seven extra bases at the 5′ and 3′ ends, respectively (underlined in figure 1A), necessary for introducing the N(t/b).BbvCI nicking sites (figure 1B). The second plasmid, pSJ3, matches M13mp2 closely over the *lacZα* sequence that actually encodes protein (the additional upstream sequences present in both M13mp2 and pSJ2 are predominantly the *lac* promoter), with minor base changes at the 5′ and 3′ termini to accommodate the N(t/b).Bpu10I nicking enzyme used in this case (figure 1B). Further very slight sequence deviations are caused by the removal a dam methylation sites (GATC) from both pSJ2 and pSJ3 (figure 1A), to eliminate any complications arsing from dam-targeted base mismatch repair [19]. With pSJ2 and pSJ3 the size of the gapped regions, ultimately used for DNA polymerase fidelity assay are, 288 and 163 bases, respectively.

Detectable sites in pSJ2 and pSJ3

The main features of pSJ2 and pSJ3 are illustrated in figure 1B. Both plasmids contain a *lacZα* gene sequence (given in full in figure 1A) flanked by two nicking endonuclease sites. The nicking enzymes (N.BbvCI and N.Bpu10I) are available as “t” or “b” forms, which specifically cut the “inner” (containing the *lacZα* coding sequence) or the “outer” strand (containing the lacZα non-coding sequence) of the plasmids shown in figure 1B. Thus the alternative use of “t” and “b” gives two orientations of the gapped plasmid, named + (inner

coding strand removed) or – (outer non-coding strand removed) in earlier studies with pSJ1 [14]. All the studies reported in this publication used the + gapped forms i.e. had the inner coding strand removed with the "t" variants of the nicking enzymes. However, a small number of investigation with the – orientations gave identical results (data not shown). A unique EcoRI restriction site is located within the *lacZα* gene, important for subsequently analysing the efficiency of the gapping process and for monitoring extension of the gapped plasmid by DNA polymerase. As pSJ2 and pSJ3 have almost identical *lacZα* sequences to M13mp2, almost all the detectable sites (i.e. changes to bases that result in an inactive the α-peptide) are known by simple comparison. The detectability of the few *lacZα* bases in pSJ2 and 3, absent in the bacteriophage, was determined using site-directed mutagenesis. Each novel base was systematically changed to all others (in the case of the *lac* promoter) or, with protein coding sequences, to any base that resulted in a change of amino acid. Only one new detectable site, for base substitutions, was found. With the underlined CGCAGCC sequence in pSJ2 (figure 1b), an A to C change gave an inactive lacZα peptide and white colonies. All other substitutions were silent. The number of detectable sites for M13mp2, pSJ2 and pSJ3 is summarised in table 1 and as expected are similar, providing the smaller size of the *lacZα* gene in pSJ3 is taken into account. A full list of the detectable sites is given in the supplementary material (figure S2).

Preparation of gapped pSJ2 and pSJ3

With pSJ1, preparation of the gapped derivative required cutting the plasmid with the appropriate nicking enzyme and heating in the presence of a single-stranded competitor DNA [14]. The competitor was designed to be complementary to the nicked strand and so to sequester it during the heat/cool cycle to give the gapped plasmid. Previously single-stranded competitor preparation used PCR amplification with two primers; one with a 5′-OH and the other containing a 5′-phosphate. Subsequent degradation of the PCR product with λ

exonuclease, an enzyme which preferentially degrades duplex DNA from ends containing a 5′-phosphate, yields the desired single strand [23]. Unfortunately λ exonuclease does not absolutely discriminate between 5′-phosphorylated and 5′-OH ends and, therefore, some destruction of the required strand occurs. To further improve selectivity the 5′-phosphorylated PCR primer has been combined with one containing a 5′-OH group and phosphorothioate modifications at the three internucleotide phosphodiester groups nearest the 5′ end (supplementary material, figure S1). DNA substituted with phosphorothioates is commonly observed to be nuclease resistant [24] and the new combination of primers is expected to better target λ exonuclease to the 5′-phosphorylated strand. Indeed the new combination of primers reliably produced higher concentrations of competitor DNA than seen when the phosphorothioates were missing. Typically ~ 17.5 μg of single stranded product was produced from 100 μl PCR reaction volumes using the triple phosphorothioated primer. In contrast ~ 7.5 μg of product was produced with a normal primer lacking phosphorothioate protection. A single-stranded competitor produced in this manner, 288 bases in length and fully complementary to the excised *lacZ* sequence, could be used to produce gapped pSJ2, as shown in figure 2A. In this figure the starting nicked plasmid and desired gapped product are poorly resolved. However, treatment with the restriction endonuclease EcoRI converts the nicked plasmid (which contains a GAATTC EcoRI site in double stranded DNA) to the linear form. The gapped plasmid, in which GAATTC is in single stranded DNA, is inert towards EcoRI. The gapped and linear plasmids are well resolved (figure 2A) and the relative intensities of their bands are indicative of the gapping yield. In agreement with earlier studies using pSJ1, a ten-fold excess of the competitor was optimal [14].

While the use of phosphorothioate-containing primers improves the PCR method for producing competitor, it is still difficult to generate large amounts of material, which is important as a ten-fold excess is required. As an alternative the use of chemically synthesised

DNA competitors has been explored. Standard solid phase synthesis of oligodeoxynucleotides is limited to lengths of around 100 bases [25], shorter than the gapped regions produced when using pSJ2 (288 bases) and pSJ3 (163 bases). However, two competitors about eighty bases in length, designed to almost completely cover the 163 nucleotide *lacZα* gene of pSJ3 (supplementary material, figure S1), were effective in gapping (figure 2B). Again EcoRI treatment was required to distinguish the starting nicked plasmid and the gapped product. These two oligodeoxynucleotides could be used without purification but a high ratio of competitor to plasmid, 50:1, was needed (with the PCR/λ exonuclease method a ten-fold excess suffices). Nevertheless the higher amounts of material available from chemical synthesis compared to PCR, make this approach worthwhile. Unfortunately, the use of chemically synthesised DNA was relatively inefficient with pSJ2. Attempts to gap the 288 nucleotide *lacZα* region with two oligodeoxynucleotide sets (the first consisting of five strands, the second four; each DNA strand between fifty and eighty bases long) were largely unsuccessful (supplementary material, figure S1). Much of the starting nicked double stranded plasmid remained in the mixture and little of the required gapped derivative was prodcued (supplementary material, figure S3). Therefore, the PCR method for competitor preparation had to be used with this plasmid.

Purification of gapped pSJ2 and pSJ3

In earlier work gapped pSJ1 was purified by gel electrophoresis, a method that limits the amount of product that can easily be prepared [14]. Gapped plasmids contain a stretch of single stranded DNA, which results in the exposure of relatively hydrophobic bases, normally buried in double-stranded DNA. This enabled their isolation using benzoylated-naphthoylated-DEAE-cellulose (BND-cellulose), a material which interacts strongly with hydrophobic regions and so binds tightly to gapped plasmids [26,27]. Elution of the gapped plasmids from BND-cellulose uses a solution of caffeine. Purification using BND-cellulose is simpler and quicker than gel electrophoresis and produces larger quantities of gapped

plasmid. Starting from 200 μg of pSJ2 or pSJ3, 80 μg of gapped plasmid (40 % yield) was typically obtained and shown by gel electrophoresis to be extremely pure (figure 2C). Treatment of the purified gapped plasmids with EcoRI did not give rise to an additional linear product of faster mobility, as would be expected in cases of contamination with nicked starting plasmid (figure 2C). The BND-cellulose purification can be scaled up to handle mg quantities of plasmid. As described below plasmids isolated using BND-cellulose resulted in a lower background mutation rate than that observed using electrophoresis, presumably arising from less damage to the DNA bases. This is a significant advantage for fidelity measurements with highly accurate DNA polymerases.

Expression frequency of pSJ2 and pSJ3

When a polymerase makes a mistake during the filling in of a gapped plasmid, either a heteroduplex with a single base mismatch or a frameshift usually results. However, this will only be scored as a polymerase error if the coding information in the newly synthesised strand is used to direct lacZα peptide synthesis. The gap-filled pSJ2 and pSJ3 plasmids used for *E.coli* transformation are dam-neutral i.e. all the GATC sites outside the *lacZα* gene are fully methylated and the gene itself lacks dam sequences. Therefore, hemi-methylated dam sequences, which would arise from polymerase filling of a *lacZα* gene containing GATC runs, are lacking. Hemi-methylation of dam sequences is used to guide mismatch repair in *E.coli* [19], and, in its absence, one would predict that base mismatches would be restored in an undirected manner. Thus, when a mismatch is produced, the polymerase introduced base and the original template base are each replaced 50 % of the time, a number that corresponds to the expression frequency. However, polymerase-filled pSJ2 and pSJ3 also contain a nick at the 3′ end of the *lacZα* gene, which could influence mismatch repair [19], making it important to experimentally determine the expression frequency. To do this a plasmid identical to that produced after a mistake in gap filling (i.e. with a detectable error in the equivalent of

polymerase-generated strand and a 3′ nick) has been generated. Full details are given in the supplementary material (figure S4) but, in brief, involve changing a single C to a T (figure 1A), converting a CAA codon (which specifies Gln13 in the lacZα peptide) to a TAA stop codon, giving a truncated inactive α-peptide. The sequences specifying an "active" and "inactive" lacZα peptide can be blended to give a heteroduplex with a G:T mismatch (supplementary material, figure S4). In this heteroduplex the non-coding strand bears the wild type triplet (TTG, specifying CAA on the coding strand following replication) and the coding strand possesses the stop codon (TAA). The coding strand additionally has a nick at the end of the *lacZα* gene and, therefore, the coding and non-coding strands are equivalent to newly synthesised and parental sequences in a polymerase extension assay. Depending on how the G:T mismatch is repaired, either an active (blue colonies) or inactive (white colonies) lacZα peptide will be produced, with the ratio of blue /white colonies reporting on which base in the mismatch is changed. When the G:T heteroduplex, was used to transform *E.coli*, 44.4 % of the colonies observed were white, resulting from a truncation in the lacZα peptide due to the stop codon (table 2). For completely random repair a value of 50 % would be expected and, therefore, the value of 44.4% indicates that repair of base mismatches in pSJ2 (and presumably pSJ3) is essentially random. For all subsequent experiments 0.444 has been used as the expression frequency. A control experiment used a homoduplex, having a wild type CAA (Gln13) codon, but produced in exactly the same manner as the heteroduplex (supplementary material, figure S4). Transformation of *E.coli* with this control plasmid resulted in only blue colonies; no white colonies, which would indicate a mutation in the *lacZα* gene, were observed (table 2).

Background mutation frequency found with pSJ2 and pSJ3

Ideally, when a plasmid or bacteriophage containing the *lacZα* gene is used to transform a complementing *E.coli* host, all colonies/plaques should be blue in the presence of X-gal, due

to the production of an active β-galactosidase. Invariably, a small number of white colonies are seen, arising from damage to the bases (some of which will inactivate the α-peptide) during the manipulations necessary to prepare and purify the gapped substrate. The resulting background mutation frequency makes it difficult to investigate highly accurate polymerases, which often produce extremely few additional errors [12, 22]. The background rates seen with BND-cellulose purified gapped pSJ2 and pSJ3 are summarised in table 3, along with that found for starting pSJ2 itself. Very few white colonies are observed, making it difficult to determine the extremely low background mutation rate accurately. For comparison the background seen using gapped pSJ2 purified by gel electrophoresis is also given, as are previous values observed for nicked pSJ1 [14] and M13mp2 [12,13]. It is clear that gapped pSJ2 purified using the BND-cellulose method shows a 5-fold reduction in background errors when compared to the gel extraction method. BND-cellulose purified gapped pSJ3 has an even lower background mutation frequency, almost certainly due to the shorter gap in *lacZα*. The very low backgrounds observed with pSJ2 and pSJ3 represents a significant improvement over earlier *lacZα*-based systems. Although only four white colonies were observed with pSJ2 and pSJ3, these have been fully sequenced to confirm the presence of a mutation. In three cases a C→T/G→A transition was observed (table 3), a mutation also overrepresented in an earlier study using pSJ1 [14]. This transition is most likely caused by the deamination of cytosine to uracil, a thymidine mimic.

Validation of pSJ2 and pSJ3

In order to check that the new plasmids are suitable for assessing the fidelity of DNA polymerases, error rates have been determined for a number of well characterised enzymes. Studies have been carried out with the Klenow fragment of the family A polymerase from *Thermus aquatics* (Taq-Pol) and three variants of the family B polymerase from *Pyrococcus furiosus* (Pfu-Pol); wild type, a 3′-5′ proof reading exonuclease deficient mutant

(D215A/E143A) [16] and an even more error prone variant (D215A/E143A/D473G) [17]. In each case the polymerase was used to fill gapped pSJ2 and pSJ3 *in vitro* and successful extension confirmed by EcoRI digestion and gel electrophoresis (figure 3). The filled plasmids were used for the transformation of *E.coli* and the numbers of blue and white colonies counted, enabling the determination of the mutation frequency (MF) (table 4). To determine the error rate (ER) the following equation, which makes explicit the reasons for determining the detectable sites (D) and expression frequency (P), was used [12, 22]:

$$ER = \frac{Ni/_{N} \times MF}{D \times P}$$

Ni = number of a particular type of mutation (usually a deletion/insertion or base substitution)

N = total number of mutations

MF = observed mutation frequency – background mutation frequency

D = number of detectable sites for a particular mutation

P = probability of expressing the mutant *lacZα* gene (expression frequency)

With the above equation, the type of mutation (N_i) can only be determined by DNA sequencing of mutant (white colonies) *lacZα* genes. In the absence of sequencing $N_i/N = 1$ and the equation can only determine total mutations [12, 13, 22].

The fidelity of wild type Pfu-Pol has been characterised previously and typically an error rate of between 1.3 x 10^{-6} and 1.6 x 10^{-6} is observed [6, 8], similar to the values observed using both pSJ2 and pSJ3 (table 4). Pfu-Pol lacking the 3′-5′ exonuclease (D215A/E143A) has an increased error rate due to loss of proofreading activity and an

approximately two to four-fold decrease in fidelity is seen with the plasmid systems. An additional mutation (D475G) to a loop in the fingers domain further increases the number of mistakes made on replication [17] and an approximately a three-fold increase in error rate is observed when compared to Pfu-Pol (D215A/E143A) (equivalent to about a ten-fold increase compared to the wild type). Finally the fidelity of a Taq-Pol, a family A DNA polymerase which lacks proof reading exonuclease activity was determined. The accuracy of Taq-Pol is strongly dependent on the reaction conditions and error rates of between 2×10^{-4} and 8×10^{-6} have been reported [6, 8, 13, 21], in agreement with the value of 1×10^{-5} found with pSJ2 and pSJ3. One study [6] showed that Taq-Pol was about six-times less accurate than Pfu-Pol, not too dissimilar to the decrease in fidelity found using pSJ2 and pSJ3 (table 4). The good general agreement found between earlier investigations and the current studies with pSJ2 and pSJ3 suggests both plasmids are suitable for determining the error rates of DNA polymerase.

The mutant (white) colonies from the pSJ2 assays of Pfu-Pol B wild type, and the exonuclease deficient variant were sequenced in order to determine further information on the mutation spectra. As was observed previously, a significant proportion of the mutants were C→T/G→A transitions, probably as a result of template strand cytosine deamination as described above for background mutation rate. This form of DNA damage was also observed in the pSJ1 and M13 methods, most likely also as a result of damage during substrate preparation [4, 12, 14]. However, as the background mutation frequency observed using pSJ2 and pSJ3 is lower, the influence of background mutations is reduced. In order to obtain rigorous data concerning the mutation spectrum of Pfu-Pol, a far higher number of mutants would obviously need to be sequenced. Nevertheless, this small data set confirms that each time a white mutant colony is observed with pSJ2, it is faithfully reflected in a DNA sequence error.

DISCUSSION

A considerably improved plasmid-based DNA polymerase fidelity assay has been described. The new plasmids, pSJ2 and pSJ3, are elaborations of an earlier version, pSJ1 [14] and contain a *lacZα* gene flanked by two nicking endonuclease sites. With pSJ2/3 both the number of detectable sites and the expression frequency have been determined, factors necessary for determining polymerase error rates and not previously available for pSJ1. The expression frequency of 44.4 % found for pSJ2 is near the theoretical maximum of 50 % expected for random repair of mismatches, suggesting high plasmid integrity with minimal base damage. The critical step in applying pSJ2 and pSJ3 is gapping which, as found with pSJ1, cannot simply be carried out using heat to remove the strand excised in the nicking step. Rather a complementary single-stranded competitor needs to be added to sequester the excised strand. This publication improves the preparation of relatively long competitors, using PCR followed by λ-exonuclease digestion, by exploiting the exonuclease resistance of phosphorothioate-protected primers. Although the phosphorothioate improvement consistently resulted in higher yields of competitor, PCR-based methods tend to give relatively small quantities of amplified material. Higher amounts of single-stranded DNA are available by direct chemical synthesis, but here the lengths that can be prepared become limiting. Thus two competitors were required to successfully gap pSJ3; unfortunately the use of four or five competitors to gap the longer pSJ2 failed. Gapping remains the most demanding part of the protocol and research is still required to enable the production of long competitors in high quantity or to develop an alternative protocol. The use of BND-cellulose to isolate gapped plasmids represents a major improvement over gel-based purification techniques. Greater amounts can prepared very straightforwardly and, importantly, the resulting gapped plasmids appear of much higher quality with little damage to the bases. This is evidenced by the very low background mutation rates seen for both pSJ2 and pSJ3 prepared using BND-cellulose, reduced approximately five-fold as compared to gel purified material. The low backgrounds seen with gapped pSJ2 (1×10^{-4}) and pSJ3 (3×10^{-5}) are extremely

advantageous when studying high-fidelity polymerases. The suitability of the new plasmids was confirmed by determining the error rates of Pfu-Pol and Taq-Pol, which gave values similar to those obtained by a number of other methods. So far pSJ2 and pSJ3 have only been used to test the fidelity of DNA polymerases *in vitro*. However, due to the compatibility of plasmids with many bacteria and eukaryotes, the *in vivo* study of DNA replication and repair should be possible, as has already been described for a plasmid-based mismatch repair activity assay [15].

ACKNOWLEDGEMENTS

This work was supported by the UK BBSRC (grant number: BB/F00687X/I). BJK is a BBSRC-supported PhD student. Excellent technical assistance was provided by Mrs. Pauline Heslop.

REFERENCES

[1] A. Johnson, M. O'Donell, (2006) Cellular DNA replicases: components and dynamics at the replication fork, Annu. Rev. Biochem. 74 (2006) 283-315.

[2] P.M.J. Burgers, Polymerase dynamics at the eukaryotic DNA replication fork. J. Biol. Chem. 284 (2009) 4041-4045.

[3] R.M. Schaaper, Base selection, proofreading, and mismatch repair during DNA replication in Escherichia coli, J. Biol. Chem. 268 (1993) 23762-23765.

[4] T.A. Kunkel, P.S. Alexander, The base substitution fidelity of eucaryotic DNA polymerases. Mispairing frequencies, site preferences, insertion preferences, and base substitution by dislocation, J. Biol. Chem. 261 (1986) 160-166.

[5] S.D. McCulloch, T.A. Kunkel, The fidelity of DNA synthesis by eukaryotic replicative and translesion synthesis polymerases, Cell Research. 18 (2008) 148-161.

[6] J. Cline, J.C. Braman, H.H. Hogrefe, PCR fidelity of Pfu DNA polymerase and other thermostable DNA polymerases, Nucleic Acids Research. 24 (1996) 3546-3551.

[7] A.R. Pavlov, N.V. Pavlova, S.A. Kozyavkin, A. I. Slesarev, Recent developments in the optimisation of thermostable DNA polymerases for efficient applications, Trends Biotechnol. 22 (2004) 253-260.

[8] K.S. Lundberg D.D. Shoemaker, M.W.W. Adams, M. Short, J.A. Sorge, E.J. Mather, High-fidelity amplification using a thermostable DNA polymerase isolated from Pyrococcus furiosus, Gene. 108 (1991) 1-6.

[9] A. Maor-Shoshani, N.B. Reuven, G. Tomer, Z. Livneh, Highly mutagenic replication by DNA polymerase V (UmuC) provides a mechanistic basis for SOS targetted mutagenesis, Proc. Natl. Acad. Sci. USA. 97 (2002) 565-570.

[10] K.A. Eckert, A. Mowery, S.E. Hills, Misalignment-mediated DNA polymerase β mutations: comparison of micro-satellite and frame-shift errors using a forward mutation assay, Biochemistry. 41 (2002) 10490-10498.

[11] S. Fujii, M. Akiyama, K. Aoki, Y. Sugaya, K. Higuchi, M. Hiraoka Y. Miki, N. Saitoh, K. Yoshiyama, K. Ihara, M. Seki, E. Ohtsubo, H. Maki, DNA replication errors produced by the replicative apparatus of Escherichia coli, J. Mol. Biol. 289 (1999) 835-850.

[12] K. Bebenek, T.A. Kunkel, Analyzing fidelity of DNA polymerases. Methods in Enzymology, 262 (1995) 217-232.

[13] K.R. Tindall, T.A. Kunkel, Fidelity of DNA synthesis by the Thermus aquaticus DNA polymerase, Biochemistry. 27 (1998) 6008-6013.

[14] S.K. Jozwiakowski, B.A. Connolly, Plasmid-based lacZalpha assay for DNA polymerase fidelity: application to archaeal family-B DNA polymerase, Nucleic Acids Ressearch. 37 (2009) e102.

[15] B. Zhou, C. Huang, J. Yang, J. Lu, Q. Dong, L.-Z. Sun, Preparation of heteroduplex enhanced green fluorescent protein plasmid for *in vivo* mismatch repair activity assay, Anal Biochem. 388 (2009) 167-169.

[16] S.J. Evans, M.J. Fogg, A. Mamone, M. Davis, L.H. Pearl, B.A. Connolly, Improving dideoxynucleotide-triphosphate utilisation by the hyper-thermophilic DNA polymerase from the archaeon Pyrococcus furiosus, Nucleic Acids Research. 28 (2000) 1059-1066.

[17] B.D. Biles, B.A. Connolly, Low-fidelity Pyrococcus furiosus DNA polymerase mutants useful in error-prone PCR, Nucleic Acids Research. 32 (2004) e176.

[18] L. Zheng, U. Baumann, J.-L. Reymond, An efficient one-step site-directed and site-saturation mutagenesis protocol, Nucleic Acids Research. 32 (2004) e115.

[19] P. Modrich, Methyl-directed DNA mismatch correction, J. Biol. Chem. 264 (1989) 6597-6660.

[20] J. Chiu, P.E. March, R. Lee, D. Tillett, Site-directed, ligase-independent mutagenesis (SLIM): a single-tube methodology approaching 100% efficiency in 4 h, Nucleic Acids Research, 32 (2004) e174.

[21] K.A. Eckert, T.A. Kunkel, High fidelity DNA synthesis by the Thermus aquaticus DNA polymerase, Nucleic Acids Research. 18 (1990) 3739-3744.

[22] J.M. Fortune, Y.I. Pavlov, C.M. Welch, E. Johansson, P.M.J. Burgers, T.A. Kunkel, Saccharomyces cerevisiae DNA polymerase delta: high fidelity for base substitutions but lower fidelity for single- and multi-base deletions, J. Biol. Chem. 280 (2005) 29980-29987.

[23] E. Thomas, A. Pingoud, P. Friedhoff, An efficient method for the preparation of long heteroduplex DNA as a substrate for mismatch repair by the *Escherichia coli* MutHLS system, Biol. Chem. 383 (2002) 1459-1462.

[24] P. Guga, M. Koziolkiewicz, Phosphorothioate nucleotides and oligonucleotides-recent progress in synthesis and applications, Chemistry and Biodiversity. 8 (2011) 1642-1681.

[25] R.A. Hughes, Miklos, A.D. Ellington, Gene synthesis: methods and applications, Methods in Enzymology. 498 (2011) 277-309.

[26] H. Wang, Hays, Simple and rapid preparation of gapped plasmid DNA for incorporation of oligomers containing specific DNA lesions, Molecular Biotech. 19 (2011) 133-140.

[27] I. Gillam, S. Millward, D. Blew, M. von Tigerstrom, E. Wemmer, G.M. Terner, The separation of soluble ribonucleic acids on benzoylated diethylaminoethylcellulose, Biochemistry. 6 (1967) 3043-3056.

FIGURE LEGENDS

Figure 1. The main features of pSJ2 and pSJ3, plasmids useful for measuring DNA polymerase fidelity. A) The *lacZα* peptide and *lac* promoter sequences of M13mp2, pSJ1, pSJ2 and pSJ3. Only the bases actually used in fidelity determination are shown; thus pSJ1 and pSJ3 contain the *lac* promoter but, because of the nicking site positions, this element is not used in fidelity assays with these two plasmids. The symbol ∗ indicates identical bases in the lacZα peptide (for all four sequences) and in the *lac* promoter (for M13mp2 and pSJ2). The underlined sequences are the extra bases at the extremities of pSJ2, necessary for introducing the nicking sites. The ATG triplet that encodes the first methionine in the lacZα peptide is shown, as is the dam-methylase site (GATC) site altered in pSJ2 and pSJ3. The C shown with the symbol ↓ is the base altered in pSJ2 to introduce a stop codon for expression frequency determination. In all cases the coding sequence of *lacZα* is given, which corresponds to the bases on the inner circle of the plasmids illustrated. B) Plasmid maps of pSJ2 and pSJ3 showing the location and orientation of the *lacZα* sequences and the nicking endonuclease sites (N(t/b)BbvCI for pSJ2) and (N(t/b)Bpu10I for pSJ3). The EcoRI site used for analytical purposes is also shown. The abbreviations dcs and ucs stand for downstream and upstream cutting sites.

Figure 2. Preparation of gapped pSJ2 and pSJ3. A) Gel electrophoretic analysis of the generation of gapped pSJ2. The nicked form (previously prepared by reaction of pSJ2 with Nt.BbvCI) is converted to the desired gapped product by heat/cool cycles with a 288 base competitor oligodeoxynucleotide (excess used indicated above the gel lanes). The nicked starting material and the gapped product are poorly resolved. However, EcoRI converts the nicked plasmid (but not the gapped) to the well separated linear form, enabling analysis of the progress of the gapping reactions. B) Gel electrophoretic analysis of the preparation of gapped pSJ3. The nicked form (previously prepared by reaction of pSJ3 with Nt.Bpu10I) can be converted to the desired gapped product using heat/cool cycles with two 80 base competitors

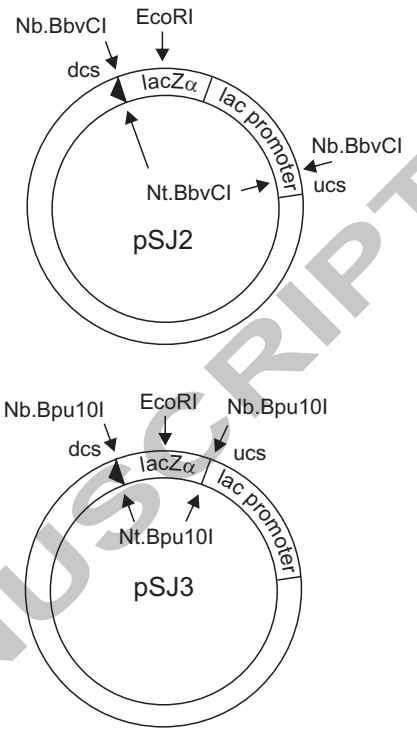
(excess used indicated above the gel lanes). The starting nicked plasmid and the gapped product are not resolved and EcoRI digestion, which converts remaining nicked pSJ3 to the linear form, but does not act on gapped form, is required for analysis. C) Gel electrophoresis of pSJ2 and pSJ3 following purification using BND-cellulose. Analyses were carried out with or without pre-treatment of EcoRI to fully control for any contaminating nicked plasmid. The size marker is a GeneRuler^{TM} 1 kb ladder (Fermantas) with the three intense bands being 1, 3 and 6 kb products as indicated on the gel.

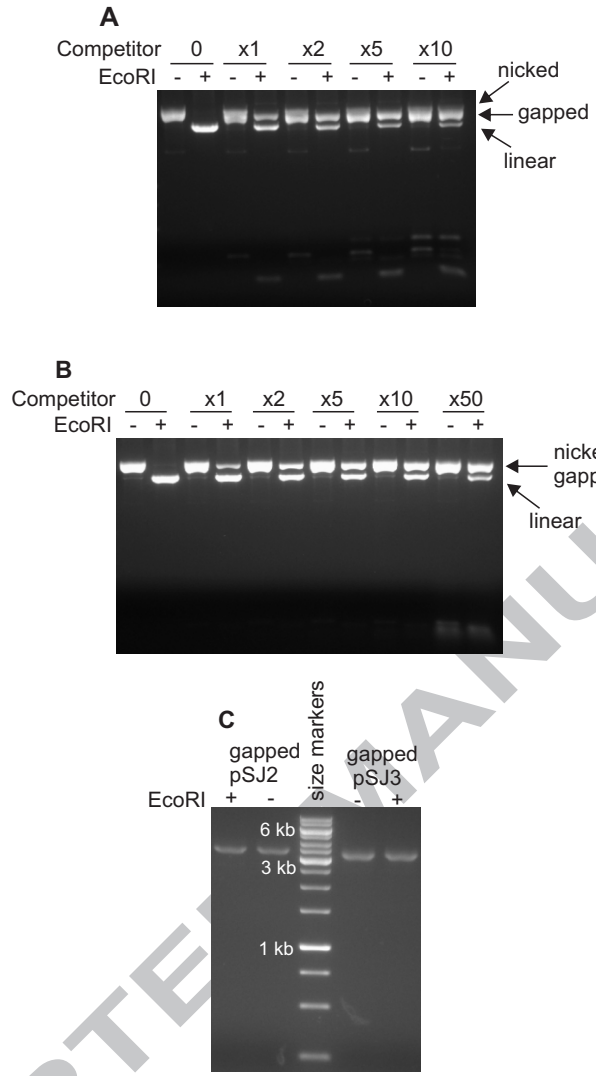
Figure 3. Filling in of gapped pSJ2 and pSJ3 with Pfu-Pol. The gel shows the starting gapped pSJ2 and pSJ3 (Pfu-Pol -, EcoRI -) and the extended product produced of addition of Pfu-Pol (Pfu-Pol +, EcoRI -). A subsequent digestion with EcoRI (Pfu-Pol +, EcoRI +) coverts the filled plasmid to the linear form, confirming extension by Pfu-Pol, which results in the EcoRI site becoming double stranded. The size marker is a GeneRuler^{TM} 1 kb ladder (Fermantas) with the three intense bands being 1, 3 and 6 kb in size as indicated on the gel.

A

```
M13mp2          ---GCGCAACGCAATTAATGTGAGTTAGCTCACTCATTAGGCACCCCAGGCTTTACACTT 57
pSJ1            ------------------------------------------------------------
pSJ2            TCAGCGCAACGCAATTAATGTGAGTTAGCTCACTCATTAGGCACCCCAGGCTTTACACTT 60
pSJ3            ------------------------------------------------------------
                ---************************************************************

M13mp2          TATGCTTCCGGCTCGTATGTTGTGTGGAATTGTGAGCGGATAACAATTTCACACAGGAAA 117
pSJ1            ------------------------------------------------------------T 1
pSJ2            TATGCTTCCGGCTCGTATGTTGTGTGGAATTGTGAGCGGATAACAATTTCACACAGGAAA 120
pSJ3            ------------------------------------------------------------T 1
                ************************************************************

                      → translation start site
M13mp2          CAGCTATGACCATGATTACGAATTC---------------------------------- 142
pSJ1            CAGCTATGACCATGATTACGAATTCGAGCTCGGTACCCGGGGATCCTCTAGAGTCGACCT 61
pSJ2            CAGCTATGACCATGATTACGAATTC---------------------------------- 145
pSJ3            CAGCTATGACCATGATTACGAATTC---------------------------------- 26
                *************************
                                        ↓
M13mp2          ------ACTGGCCGTCGTTTTACAACGTCGTGACTGGGAAAACCCTGGCGTTACCCAACT 196
pSJ1            GCAGGCACTGGCCGTCGTTTTACAACGTCGTGACTGGGAAAACCCTGGCGTTACCCAACT 121
pSJ2            ------ACTGGCCGTCGTTTTACAACGTCGTGACTGGGAAAACCCTGGCGTTACCCAACT 199
pSJ3            ------ACTGGCCGTCGTTTTACAACGTCGTGACTGGGAAAACCCTGGCGTTACCCAACT 80
                      ************************************************** 

M13mp2          TAATCGCCTTGCAGCACATCCCCCTTTCGCCAGCTGGCGTAATAGCGAAGAGGCCCGCAC 256
pSJ1            TAATCGCCTTGCAGCACATCCCCCTTTCGCCAGCTGGCGTAATAGCGAAGAGGCCCGCAC 181
pSJ2            TAATCGCCTTGCAGCACATCCCCCTTTCGCCAGCTGGCGTAATAGCGAAGAGGCCCGCAC 259
pSJ3            TAATCGCCTTGCAGCACATCCCCCTTTCGCCAGCTGGCGTAATAGCGAAGAGGCCCGCAC 140
                ************************************************************
                dam
M13mp2          CGATCGCCCTTCCCAACAGCTG---------------------------------- 278
pSJ1            CGATCGCCCTTCCCAACAGTTGCGCAGCCTGAATGGCGAATGGCGCCTGATGCGGTATTT 241
pSJ2            CGACCGCCCTTCCCAACAGCTGCGCAGCC------------------------------ 288
pSJ3            CGGTCGCCCTTCCCAACAGTTCC---------------------------------- 163
                *  *************** * ———————

M13mp2          ------------------------------------------------------------
pSJ1            TCTCCTTACGCATCTGTGCGGTATTTCACACCGCATATGGTGCACTCTCAGTACAATCTG 301
pSJ2            ------------------------------------------------------------
pSJ3            ------------------------------------------------------------

M13mp2          ------------------------
pSJ1            CTCTGATGCCGCATAGTTAAGCC 324
pSJ2            ------------------------
pSJ3            ------------------------
```

B





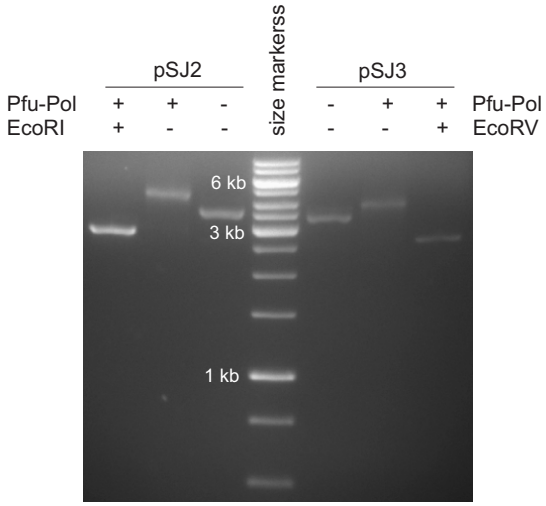


Table 1. Detectable sites within the *lacZα* gene

lacZα gene source	Length of *lacZα* gene analysed (bases)[1]	Number of detectable base substitutions	Number of detectable insertions and/or deletions	Total number of detectable sites
M13mp2[2]	278	241	199	440
pSJ2[3]	288	242	206	448
pSJ3[3]	163	166	163	329

[1]The sequences of the lacZα genes analysed are given in figure 1A

[2]Taken from work reported by Kunkel and co-workers (12, 22)

[3]It has been assumed that the bases shared between pSJ2/pSJ3 and M13mp2 have the same "detectability". The properties of the extra bases present in pSJ2 were determined by site-directed mutagenesis.

Table 2. Expression frequency of lacZα gene in pSJ2

Nicked plasmid	Total number of colonies[1]	Number of mutant (white) colonies	% mutant colonies	Expression frequency of *lacZα* gene
Heteroduplex (G:T mismatch)	7603	3379	44.4	0.444
Homoduplex (G:C base pair)	7393	0	0	Not applicable

[1]Obtained by summing three independent determinations.

Table 3. Background mutation frequencies of pSJ2 and pSJ3

Plasmid	Total number of colonies	Number of mutant (white) colonies[1]	Mutation rate (%)	Mutation frequency[2]
pSJ2[3]	39,119	1	0.0026	2.6×10^{-5}
pSJ2 (gapped, BND-cellulose purified)[3]	28,406	3	0.011	1.1×10^{-4}
pSJ2 (gapped, gel purified)[3]	26,313	14	0.053	5.3×10^{-4}
pSJ3 (gapped, BND-cellulose purified)[3]	32,519	1	0.0031	3.1×10^{-5}
pSJ1[4]	-	-	0.08	8.0×10^{-4}
M13mp2 (phage)[5]	-	-	0.04-0.06	$4.0\text{-}6.0 \times 10^{-4}$

[1]Plasmid DNA isolated from the four mutant colonies found with gapped pSJ2 and gapped pSJ3 (purified using BND-cellulose) was sequenced. Three of the mutations in the lacZα gene were C→T/G→A transitions. The forth was a T→C/A→G transition.

[2]Number of errors made per base

[3]Each data set with pSJ2 and pSJ3 was the sum of five independent experiments carried out in triplicate.

[4]Data taken from a previous study (14)

[5]Data taken from previous studies (12, 22)

Table 4. Error rates of DNA polymerases determined using pSJ2 and pSJ3

Polymerase	Gapped plasmid[1]	Number of colonies[2]	Number of mutant (white) colonies	Mutation frequency[3]	Error rate[4]
Pfu-Pol B	pSJ2	25,700	11	3.2×10^{-4}	1.6×10^{-6}
	pSJ3	20,116	11	5.2×10^{-4}	3.5×10^{-6}
Pfu-Pol B (D215A)[5]	pSJ2	14,601	20	1.3×10^{-3}	6.3×10^{-6}
	pSJ3	24,766	25	1.0×10^{-3}	6.7×10^{-6}
Pfu-Pol B (D215A/D473G)[6]	pSJ2	78,431	296	3.7×10^{-3}	1.8×10^{-5}
	pSJ3	38,836	141	3.6×10^{-3}	2.4×10^{-5}
Taq-Pol	pSJ2	46,239	98	2.0×10^{-3}	1.0×10^{-5}
	pSJ3	20,756	34	1.6×10^{-3}	1.1×10^{-5}

[1]All of the gapped plasmids used in these experiments had the coding strand (inner strand in figure 1B) removed by treatment with Nt.BbvC1 (pSJ2) or Nt.Bpu10I (pSJ3).

[2]Sum of three independent experiments, each consisting of five repeats.

[3]The mutation frequencies given here have had the background mutation frequencies found for gapped pSJ2 and pSJ3 subtracted.

[4]Error rates were calculated using the formula given in the text. An expression frequency (P) of 0.444 was used. In the absence of extensive DNA sequencing an N_i/N value of 1 was used and the number of detectable sites (D) was the sum of the values for base substitutions plus insertions/deletions i.e. 448 for pSJ2 and 329 for pSJ3 (table 1).

[5]The mutation D215A disables the 3′-5′ proof reading exonuclease activity (16)

[6]The double mutation D215A/D473G has even lower fidelity compared than the exo$^-$ single mutant D215A (17).